DOCUMENT RESUME

ED 279 665 TM 870 052

AUTHOR Carlson, Ken
TITLE The National Assessment of Educational Progress in Social Studies.
PUB DATE Aug 86
NOTE 22p.; One of 46 papers commissioned by the Study Group on the National Assessment of Student Achievement and cited in Appendix B to their final report "The Nation's Report Card" (TM 870 049). For other papers in this group, see TM 870 050-094.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Citizenship Education; Content Analysis; Content Validity; Educational Assessment; *Educational Objectives; Educational Testing; Elementary Secondary Education; Multiple Choice Tests; *National Surveys; *Social Studies; Test Construction; *Testing Problems; *Testing Programs; Test Length; United States History; Young Adults
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

This paper discusses the content of the social studies tests of the 1981-82 National Assessment of Educational Progress (NAEP). Selected social studies and citizenship items were administered to 3,200 students aged 9, 13, and 17 and to adults aged 26-35. Twelve sources describing the social studies tests were reviewed, particularly the Citizenship and Social Studies Objectives and the Released Exercises, containing about one-fourth of the test items. The objectives were developed by 255 people, and resulted in 196 specific objectives in five major categories, demonstrating: (1) skills necessary to acquire information; (2) skills necessary to use information; (3) understanding of individual development and communication skills; (4) understanding of and interest in the ways humans organize, adapt to, and change their environment; and (5) understanding of and interest in the development of the United States. There were also 32 mid-level objectives, which are appended. Most of the objectives dealt with content, while some covered discourse, thinking, and values and attitudes. There was concern about content validity, since the categories were not evenly represented. In a number of cases, an argument could be made about the correct alternative from the multiple choice items. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED279665

The National Assessment of Educational Progress in Social Studies

Ken Carlson

Rutgers University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. Carlson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper commissioned by

THE STUDY GROUP ON THE NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENT

1986

BEST COPY AVAILABLE

TM 870 052

THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS IN SOCIAL STUDIES

Ken Carlson
Rutgers University

August 1986

Introduction

The history of NAEP assessment in social studies has been short and recent, even though the U.S. Office of Education was charged at its chartering in 1867 with determining the nation's progress in education.

> In the 1981-82 school year, the National Assessment of Educational Progress will conduct a third survey of the citizenship and social studies achievement of young people across the nation. Citizenship was first assessed in 1969-70; social studies was assessed for the first time in 1971-72; both were reassessed in 1975-76. Because there is a high degree of similarity between the goals considered important in both citizenship and social studies education, National Assessment has combined the two areas into one comprehensive area for the 1981-82 assessment. (Citizenship and Social Studies Objectives, 1981-82, p. 1)

The assessment is administered to 9-, 13-, and 17-year olds, as well as young adults ages 26-35. The first three age groups tend to be in grades 4, 8 and 11. Some test items are administered to only one of the school age groups, some to two of the groups, and still other items are administered to all three groups.

The 1981-82 assessment was administered to probability samples of

about 2,000 individuals at age 9 and 1,200 at ages 13 and 17. It is this latest assessment which will be the focus of this paper since the latest assessment is likely to be the best predictor of the next assessment.

Sources

The sources that were obtained from the NAEP offices at Educational Testing Service in Princeton, New Jersey, are:

Code	Title	Cost
07-C-00	Citizenship: An Overview	6.00
07-CS-01	Education for Citizenship	4.00
07-CS-02	Changes in Political Knowledge	5.00
07-CS-21	Technical Documentation for Performance	21.00
07-CS-25	Citizenship/Social Studies Released Set	35.00
07-CS-26	Technical Appendix to Released Set	8.00
13-CS-01	Achievement of Young Americans	3.00
13-CS-25	Citizenship/Social Studies Released Set, 1981-82	28.00
13-MCS-40	Procedural Handbook, 1981	13.00
no code	Social Studies Objectives, 1971-72	free
no code	Citizenship Objectives, 1974-75	free
no code	Citizenship and Social Studies Objectives, 1981-82	free

NAEP is to be commended for making available so much information with which to analyze its assessment program, and for making the information available to the researcher free of charge.

Consistent with the focus of this paper, the two sources which were reviewed most thoroughly are the 1981-82 Citizenship and Social Studies Objectives, and the 1981-82 Released Exercises.

The Objectives

The first order of business in setting objectives for the NAEP social studies assessment is picking the people who will be asked what the important objectives of social studies education are. Given the diverse nature of the social studies, it is impossible to have a list of consultants that is both short and representative. The list was long for the earlier assessments, and it is even longer for the 1981-82 assessment.

For the 1981-82 assessment, there were 255 people who participated in the development of the objectives. Included among these people was the chairman of the committee for which this paper is being prepared, Jack Fraenkel. Organizational representation was from state education departments, bar associations, high schools (public and private), state legislatures, scholarly associations, advocacy groups, the military, industry, PTAs, labor unions, ethnic groups, law enforcement agencies, the NEA and the AFT, farm bureaus, the media, Leagues of Women Voters, and academia.

Such broad representation has the virtue of inclusiveness. It may, however, yield the vice of proliferation in terms of the objectives that get developed. (Again, however, given the comprehensiveness of the social studies, a proliferation of objectives could occur with a single consultant!) For the 1981-82 assessment there are 196 operational objectives. This is a conservative count since many of the objectives are

compound ones. The 196 specific objectives are gathered under five major headings:

I. Demonstrates Skills Necessary to Acquire Information

II. Demonstrates Skills Necessary to Use Information

III. Demonstrates an Understanding of Individual Development and the Skills Necessary to Communicate With Others

IV. Demonstrates an Understanding of and Interest in the Ways Human Beings Organize, Adapt to and Change Their Environments

V. Demonstrates an Understanding of and Interest in the Development of the United States.

Between the major headings and the specific objectives are 32 mid-level objectives. These are presented in the Appendix since they provide sufficient specificity for the reader to know with some exactness what is t ing sought and yet they are not numbingly numerous. Furthermore, the released exercises which will be discussed later are keyed to the mid-level objectives.

For the 1974-75 assessment there were 25 objectives for the social studies section and 36 objectives for the citizenship objectives, or a total of 61 objectives. That is considerably fewer than the 196 objectives in 1981-82. It is possible to argue that because the two sets of objectives are couched somewhat differently, a surface comparison cannot be made. Even if that argument has weight, the important concern to be registered here is whether it is possible to measure either number of objectives with a 45-minute test. Some of the objectives may not get

measured at all, and others may be measured so superficially that the results have no validity.

To illustrate the concern, we can use one of the mid-level objectives from the 1981-82 assessment. This objective reads "Communicates in graphic and oral forms." Since the assessment consists of a paper-and-pencil test, it would seem that the second part of the objective (oral communication) is not being measured. It is, of course, possible to ask students to check yes or no as to whether they communicate orally, but that would trivialize the objective beyond defense.

Further elaboration on this concern, as well as the presentation of other concerns, is undertaken in the discussion of the released exercises from the 1981-82 assessment. Before beginning that discussion, however, an attempt will be made to separate the 196 detailed objectives into the four categories of concern to the committee: discourse, content, thinking, and values and attitudes.

Some of the objectives, because of their compound nature, fall into more than one category. For example, the objective "Creating, analyzing and evaluating alternative futures for the world (e.g., predicting what future value systems or lifestyles might be like; predicting the form of future economic, political or social orders; predicting the uses of technology in the future; projecting population trends)" would seem to include all four categories. With the objectives that are directed primarily to thinking skills, a case can be made that the thinking skills are essential to effective discourse and therefore the objectives overlap the thinking and discourse categories. This case is buttressed by the fact

that the released exercises for such objectives call for open-ended responses. Even so, the decision made by the writer was that if a thinking objective could be measured with a multiple choice item, then it was not to be considered as including the discourse category.

Table 1 presents the breakdown as the writer finally agreed to it with himself.

TABLE 1

NUMBER OF OBJECTIVES FOR EACH CATEGORY

Discourse	9
Content	120
Thinking	32
Values and Attitudes	26
Discourse and Thinking	6
Thinking and Values	1
Content and Values	1
Discourse, Content, Thinking and Values	1
TOTAL	196

Since the content objectives are so predominant, something further should be said about them. By and large, they are of the "big ideas" type. Understanding signal aspects of the American political and legal systems, individual rights, economics, labor-management relations, environmental management, etc., are objectives with which it is difficult to quarrel. Moreover, the understandings sought through the

content objectives pretty well cover the several social studies disciplines and probably reflect well the concerns of the organizations that participated in the development of the objectives.

While understandably not as numerous as the content objectives, the thinking objectives cover well the hierarchy of cognitive skills, from acquiring information through the senses to the uses of logic, the testing of hypotheses, and the synthesizing of ideas.

In conclusion, one might paraphrase Ed McMahon's frequent assurance and say that the NAEP social studies objectives cover aaaall that anyone should want to learn in social studies, Johnny. With 255 people suggesting the objectives, it would have been surprising if anything of substance had been left out. Thus, the concern is not with the quality of the objectives but rather with their sheer quantity.

Released Exercises

The first thing to note here is that only about one-fourth of the exercises from the 1981-82 assessment were released to the public. The remainder were held back for future administration to determine changes in the performance level of young Americans. The basis upon which an exercise was chosen for release or retention is not given in the printed material from NAEP. However, a phone conversation with Archie Lapointe, executive director of the NAEP, on July 30, revealed that the retained exercises are those that "perform best statistically" and have durability in terms of what social studies will continue to emphasize. There has also been a recent decision by the NAEP Policy Board to allow

for full release of the exercises to school districts, state education agencies, and other entities which have a legitimate interest and which sign an agreement to keep the exercises secure. This writer was unaware of the opportunity to see all the exercises at the time he picked up the materials from the NAEP offices in Princeton. Thus, the following discussion is based only on the released exercises he was given, which exercises are now assumed to be those which did not "perform well statistically."

A second thing to note is that each released exercise is identified by the objective to which it corresponds. However, the identification is only to the mid-level objectives (see Appendix). Altogether, 57 exercises were released from the 1981-82 assessment. (NAEP claims there were 61 released exercises, but repeated cross checking by this writer failed to produce more than 57.)

Table 2 shows how well the released exercises represent the mid-level objectives. Twenty-five objectives are covered; seven are not. Thirteen objectives are represented by a single exercise each, whereas one objective has nine exercises.

Interestingly, the objective referred to earlier as inherently unmeasurable with a paper-and-pencil test (the objective on communicating in graphic and oral form) has no released exercise. Without belaboring the point, it is worth noting that two of the five major headings are very poorly represented among the released exercises. These are "Demonstrates an Understanding of Individual Development and the Skills Necessary to Communicate With Others" and "Demonstrates an Understanding of and Interest in the Ways Human Beings Organize, Adapt to and Change Their

Environments." One of the headings is very well represented. This is "Demonstrates an Understanding of and Interest in the Development of the United States." If the released exercises are at all representative of the total distribution of exercises, it can be argued that the assessment has content validity insofar as it tests most for what is most emphasized in the curriculum. This category is also the most content-heavy of the five major categories, a heaviness which, again, is typical of social studies curricula.

TABLE 2

NUMBER OF RELEASED EXERCISES PER MID-LEVEL OBJECTIVE*

Objective	No. of Exercises	Objective	No. of Exercises
I.A.		IV.A.	
B.	3	B.	1
C.	2	C.	1
II.A.	2	D.	
B.	1	E.	
C.	3	F.	1
D.	1	G.	1
III.A.	1	H.	3
B.		V.A.	1
C.		B.	1
D.		C.	2
E.	1	D.	5

Objective	No. of Exercises
III.F.	1

Objective	No. of Exercises
V.E.	3
F.	5
G.	1
H.	4
I.	2
J.	1
K.	9

*See Appendix for wording of the objectives.

Before discussing the quality of the released exercises themselves, there is an interesting anomaly to be noted. The introductory material to the released exercises includes a brief description of the scoring procedures for the open-ended exercises. (The description in the Procedural Handbook is not much longer.) The description tells how some of the open-ended exercises were scored independently by two scorers so that some index of inter-rater reliability could be computed. The percentages of scorer agreement are said to range from 79.9 percent to 98.8 percent, with most scorer agreements greater than 90 percent. The reader is referred to Exhibit 2 for a display of the numbers. Strangely, Exhibit 2 does not have anything to do with scorer agreement on the social studies exercises. It displays the results for the scoring agreement on the mathematics exercises, an area where it would seem easier to achieve such agreement.

There are six concerns raised by the released exercises. In no particular order, the concerns are:

1. the acceptance of very simple answers
2. the rejection of answers that seem to be as good as some of the accepted answers
3. the arguability of the officially correct answers to some of the multiple choice questions
4. questions for which desired answers are obvious
5. questions which don't really tap the objectives they are supposed to
6. hostile and cynical answers and the response to them.

An example of simple answers which were scored as satisfactory are the answers to the question "Name as many reasons as you can why laws are needed." Answers such as "to help make a better life" (age 13), "to better society" (age 17), and "so people will do things right" (age 17) were deemed acceptable. For the question "If a person thinks the Governor or President is doing a <u>good</u> job, is it all right for the person to tell other people about it?" some of the accepted yes explanations are no more than declarative paraphrases of the question itself. "If a person thinks he is doing a good job they have a right to say so" (age 13). "Yes because people should be able to say what they want to" (age 13). As a last example (although far from the last possible one), there were answers to the question "What would you do if your community made a new law which you thought was very unfair?" "Try somehow to get rid of this law" (age 17). "I'd probably have something to say about it - express my opinion" (age 17).

More worrisome than the leniency with which simple answers are accepted is the arbitrariness with which some equally good answers are rejected. The question "Are there any kinds of people you think should not be permitted to get together to make plans or do things?" elicited such accepted answers as "chronic trouble makers" (age 13), "people who are trying to overthrow the government" (age 13), and "some people get together to cause riots" (age 13), and such rejected answers as "Negroes when they want to riot" (age 13). In several instances a rejected answer is substantively indistinguishable from an accepted one. Again, the question "What would you do if your community made a law which you thought was very unfair?" An accepted answer was "I'd probably try to change it in some way, if it affected me very much" (age 17). A rejected answer was "Try to make the law the way I feel it should be" (age 17). Often acceptance or rejection of an answer appears to be based on the scorer's ideology. For example, the question "Do you think you can have any influence on decisions of the national government?" yielded no accepted negative answers. All those who answered that they could have an influence by voting were scored correct, but the one person who said that he or she could not have an influence because not a citizen and therefore ineligible to vote was scored incorrect. The student who answered "I am not old enough to vote and right now I am just a very small cog in the machinery" (age 17) gave as cogent an explanation for his/her no answer as can be found for many of the yes answers. But negative answers were not tolerated, apparently because they were thought to be violative of the belief in individual citizen efficacy which schools are expected to foster.

As examples of the arguability of the officially correct answers to some of the multiple choice exercises, there are the following. One question asked for pupils to check whether a given action would or would not help a small group of students work well together in planning a school festival. "One person would be the leader all the time" and "We would ask the teacher what to do" were both actions which were officially rejected as being helpful to the group. Another question asked how Linda could make friends with a new girl in her class if the new girl speaks French and knows very little English. The officially correct answer was that Linda should sit with the new girl at lunch and try to learn a few words of French. Officially incorrect answers were that Linda talk to the new girl only in English and hope she understands, and that Linda learn French before trying to talk to the new girl. A third example is the question which requires students to mark true or false that, once a law is passed by the U.S. Congress, an agency that enforces the law can sometimes specify what the law means in specific cases. The official answer here is false, enforcement agencies have no specification authority. As can be seen, the answer choices in these examples are worded so ambiguously that a wrong answer could be correct if a student had a chance to explain his understanding of the wording. (On the day this is being typed there appears in The New York Times an obituary for Banesh Hoffman, who used to rail against the very thing being described here.)

For several exercises the desired answers are obvious. One question begins "During an election campaign" and is followed by several activities the student would like to do. It is obvious that students

should like to do all the activities, and a student who indicates honestly that he/she would not like to engage in one of these activities gets marked wrong. Another question asks students to check from a list of citizenship activities all those in which they have personally engaged, e.g., taking part in a demonstration. For a student to check that a certain activity is one in which he/she has not taken part is to be marked wrong. Finally, there is a yes/no question which simply asks students whether they will vote when they are 18. "No" is obviously the wrong answer. It may be replied that there is no inhibition of honest answers and no inducement to dishonesty with these questions since student performance on the NAEP has neither personal nor institutional consequences. That situation would change if it should come to pass that state-by-state, district-by-district, or school-by-school comparisons are made. To do the first of these comparisons would generate pressure for the other two.

Several of the exercises do not seem to tap the objective that they are purported to measure. For example, students are asked whether it is true for them that "A lot of times I am noisy when other students are trying to say something" and "When other students are talking, I usually wait for them to finish before I try to say something." This exercise is supposed to plumb objective IV.H - students' commitment to human rights worldwide. The exercise that calls on students to give three reasons why it is important to have police is supposed to tap into objective V.E - an understanding of the basis and organization of the legal system in the United States. There is a multiple choice question on which the student is to pick the federal tax which yields the

greatest amount of revenue. A correct answer to this question is presumed to show that the student understands economics in the United States - objective V.I. A correct answer to the question of which geographic feature covers the greatest area of the globe - mountains, plains, lakes, or oceans - indicates that the respondent understands the relationship between people and the natural environment - objective IV.F.

Finally, there are a lot of very hostile and cynical answers. There are answers which show gross contempt for the police and the legal system generally. There are answers which are defiantly racist. There are answers which advocate beatings and shootings as ways of "persuading" people. The content of these answers certainly makes them troublesome, but the students' motivations in writing the answers are unknown. Unlike the previous concerns, the concern here is not about the quality of the exercises and the scoring but rather about the uses to which the test results are put. If students are giving answers which are indisputably at odds with the objectives of social studies education, it would be well to have that information fed back to teachers with some advice on how to handle the situation.

Summary

Major concerns are raised by a review of the NAEP assessment in social studies. The first concern is the sheer number of assessment objectives. It is recognized that social studies is the most comprehensive school subject, and that this feature calls for a wide array of input into the development of the test objectives, and that these combined phenomena then result in a slew of objectives. Still, if the assessment is

to be kept to a 45-minute paper-and-pencil test for logistical reasons, there may be only public relations value to having more objectives than can be measured or measured validly.

The second concern is particular to the committee for which this paper was prepared. The committee thinks that "discourse" is an important category of objectives. Among the 196 NAEP objectives, however, there are only 15 which this writer could construe as falling clearly into the category of discourse. In looking at the released exercises which are open-ended and therefore might presumably be used to assess discourse, one is struck by the limited space allowed for answers and by the brevity of the answers students give. One is also struck by the leniency with which truistic answers are accepted, as well as by the arbitrariness (or ideological bias) with which equally good answers are rejected.

The released exercises which are of the multiple choice type arouse other concerns. The officially correct answers to some of these exercises are certainly debatable, and mostly because the answer choices are worded so ambiguously. This is the traditional complaint about multiple choice items, and the fact that the NAEP has not succeeded in breaking with tradition is disappointing.

Other multiple choice questions ask students how they have behaved or would behave in certain circumstances. The value loading to these questions is so transparent that any student who wants to add several quick points to his or her test score can do so easily.

There are also questions for which any resemblance to the objective being measured is barely perceptible. Lofty objectives get

trivialized by questions which tap into them in the most rudimentary fashion.

The last concern is with the uses to which the test results are put. When a student writes that his reaction to a waiter who refuses to seat a minority couple in a restaurant would be to congratulate the waiter and give him an extra big tip, one very much desires a chance to talk to the student. It is not clear, however, if this answer is communicated to the student's teacher or just buried in the national aggregation. To know that some students express reprehensible values is not very useful information if nothing is done with it.

APPENDIX

MAJOR HEADINGS AND MID-LEVEL OBJECTIVES OF THE 1981-82 CITIZENSHIP/SOCIAL STUDIES ASSESSMENT

I. Demonstrates Skills Necessary to Acquire Information

- A. Uses the senses.
- B. Uses sources such as card catalogues and indexes, case studies, computers, drawings, films, globes and other models, graphs, maps, newspapers, photos, pictures, radio, recordings, reference books, slides, tapes, television.
- C. Uses techniques such as personal interviews, written essays, polls and questionnaires.

II. Demonstrates Skills Necessary to Use Information

- A. Organizes information.
- B. Applies information.
- C. Makes decisions and solves problems.
- D. Critically evaluates information.

III. Demonstrates an Understanding of Individual Development and the Skills Necessary to Communicate with Others

- A. Examines individual beliefs, values and behaviors.
- B. Demonstrates individual development.
- C. Communicates in graphic and oral forms.
- D. Gives attention and responds to the expression of others.
- E. Interacts in groups in various capacities.
- F. Has effective relations with people having different cultural perspectives.

IV. Demonstrates an Understanding of an Interest in the Ways Human Beings Organize, Adapt to and Change Their Environments

- A. Understands the forces that shape individual human beings.

B. Understands the interrelatedness of human societies.

C. Understands the organization of human societies.

D. Understands the relationships between individuals and groups.

E. Understands the relationships among groups.

F. Understands the relationships between people and the natural environment.

G. Has an awareness of global concerns.

H. Has a commitment to human rights worldwide.

V. Demonstrates an Understanding of an Interest in the Development of the Unites States

A. Understands the principles and purposes of the United States.

B. Understands the organization and operation of the governments in the United States.

C. Understands political decision making in the United States.

D. Understands the electoral processes in the United States.

E. Understands the basis and organization of the legal system in the United States.

F. Knows rights of individuals in the United States.

G. Recognizes civil and criminal judicial systems in the United States.

H. Has a commitment to support justice and rights of all individuals.

I. Understands economics in the United States.

J. Understands major social changes that have occurred in American society.

K. Has a commitment to participating in community service and civic improvement.